HIGHER MOMENTS OF THE CLAIMS DEVELOPMENT RESULT IN GENERAL INSURANCE

BY

ROBERT SALZMANN, MARIO V. WÜTHRICH AND MICHAEL MERZ

ABSTRACT

The claims development result (CDR) is one of the major risk drivers in the profit and loss statement of a general insurance company. Therefore, the CDR has become a central object of interest under new solvency regulation. In current practice, simple methods based on the first two moments of the CDR are implemented to find a proxy for the distribution of the CDR. Such approximations based on the first two moments are rather rough and may fail to appropriately describe the shape of the distribution of the CDR. In this paper we provide an analysis of higher moments of the CDR. Within a Bayes chain ladder framework we consider two different models for which it is possible to derive analytical solutions for the higher moments of the CDR. Based on higher moments we can e.g. calculate the skewness and the excess kurtosis of the distribution of the CDR and obtain refined approximations. Moreover, a case study investigates and answers questions raised in IASB [9].

INTRODUCTION

One of the most important financial positions in the balance sheet of a general insurance company are the claims reserves. The claims reserves are used for paying the outstanding loss liability cash flows. The changes (over time) in these claims reserves, called the claims development result (CDR), are one of the major risk drivers in the profit and loss statement of a general insurance company. Therefore, in new solvency regulations, such as Solvency II, the CDR has become a central object of interest for the assessment of the reserving risk. As proposed in Solvency II and applied in current practice, a proxy for the distribution of the CDR is obtained by fitting a shifted lognormal distribution to the estimates of the first two moments of the CDR. This rather rough approach may fail to describe the shape of the distribution of the CDR. In fact, fitting a shifted lognormal distribution to the first two moments or applying bootstrapping techniques often lead to rather symmetric approximations to the distribution of the CDR which, in general, fail to fit the skewness and the excess kurtosis of the distribution of the CDR. Knowledge about the shape

Astin Bulletin 42(1), 355-384. doi: 10.2143/AST.42.1.2160747

of the distribution is important for the risk assessment of future cash flows as discussed in IASB [9]. As for instance stated in paragraph B76 of IASB [9], for the selection of the confidence level for estimating risk adjustments the insurer must take into account additional factors, such as the skewness of the probability distribution. Moreover, when it comes to selecting an appropriate technique to estimate risk adjustments, shape parameters such as the skewness have to be taken into account as stated in paragraph B95 of IASB [9]. The consideration of higher moments allows insurance companies to gain more insight into the shape of the distribution of the CDR which is essential for the application of techniques for estimating risk adjustments as discussed in IASB [9].

We work within a chain ladder (CL) framework. Since the CDR in the next accounting year results from an information update, such problems are most naturally understood recursively (see Bühlmann-Gisler [2], Section 9.2). A unified approach to incorporate new information into a CL model is provided by the Bayes CL framework (see Bühlmann et al. [1]). In this paper we consider two different distributional Bayes CL models. We emphasize that for the analysis of higher moments we need to make explicit distributional model assumptions.

As an important property of the Bayesian setup we highlight that it does not only allow for the incorporation of the information update but it also accounts for the parameter estimation uncertainty in a natural way. Moreover, our distributional model assumptions allow for the calculation of higher moments of the CDR in closed form w.r.t. the data available.

In a case study we fit the so called 4-parameter Johnson family of distributions (see Johnson et al. [11]) to moments up to order four of the CDR, i.e. in addition to the first two moments we fit the skewness and excess kurtosis of the CDR. As a special case the 4-parameter Johnson family of distributions includes the shifted lognormal distribution which is proposed in Solvency II as a proxy for the distribution of the CDR. This allows to directly compare the distributional approximation for the shifted lognormal distribution based on the first two moments with the refined approximation for the 4-parameter Johnson family of distributions based on higher moments up to order four. The case study shows that considering the shape of the distribution of the CDR significantly influences the uncertainty analysis of the CDR and answers questions raised in IASB [9].

Organization of the paper

In the following section we introduce the CDR. In Section 2 we define two different Bayes CL models. Then we calculate posterior distributions (of the underlying model parameters) which enable us to compute higher moments of the CDR in closed form (as far as they exist) for single accident years as well as for the aggregated case, see Section 3. The estimation of the structural parameters (which are needed to calibrate the models) is described in Section 4. Finally, Section 5 presents a case study and a discussion of the results. The main results are proved in the appendix.

1. Claims Development Result

In the sequel we work with claims development triangles (trapezoids) for which the single entries correspond to cumulative claims denoted by $C_{i,j} > 0$. The index $i \in \{0, \ldots, I\}$ refers to *accident years* and the index $j \in \{0, \ldots, J\}$ to *development years* ($I \geq J$). We assume that the ultimate claim for accident year $i \in \{0, \ldots, I\}$ is given by $C_{i,J}$. This means we assume that there is no development after development year J.

At time $I + t$, $t \geq 0$, we have observations denoted by

$$\mathcal{D}_{I+t} = \left\{C_{i,j};\, i + j \leq I + t,\ 0 \leq i \leq I,\ 0 \leq j \leq J\right\}.$$

This corresponds to the runoff situation because no new accident years i are added to $\mathcal{D}_{I+t}$ after year I. The period $(t-1, t]$ is called *accounting year t* and $\mathcal{D}_{I+t}$ is the information available at the end of accounting year t.

Claims reserving is basically a prediction problem and we are mostly interested in the prediction of the ultimate claim $C_{i,J}$. The Bayesian predictor for the ultimate claim $C_{i,J}$, given $\mathcal{D}_{I+t}$, is defined by

$$\widehat{C}_{i,J}^{(t)} = \mathbb{E}\left[C_{i,J} \middle| \mathcal{D}_{I+t}\right]. \tag{1.1}$$

Note that this predictor is unbiased and has minimal (conditional) L^2-uncertainty among all $\mathcal{D}_{I+t}$-measurable predictors and is therefore often called "best-estimate" predictor.

The difference in successive ultimate claim predictions resulting from an information update $\mathcal{D}_I \to \mathcal{D}_{I+1}$, see Figure 1, determines the CDR in the first accounting year $(0,1]$.

Definition 1.1. *The CDR for accident year $i = I - J + 1, \ldots, I$ in the first accounting year is defined by*

$$\mathrm{CDR}_i = \widehat{C}_{i,J}^{(0)} - \widehat{C}_{i,J}^{(1)}.$$

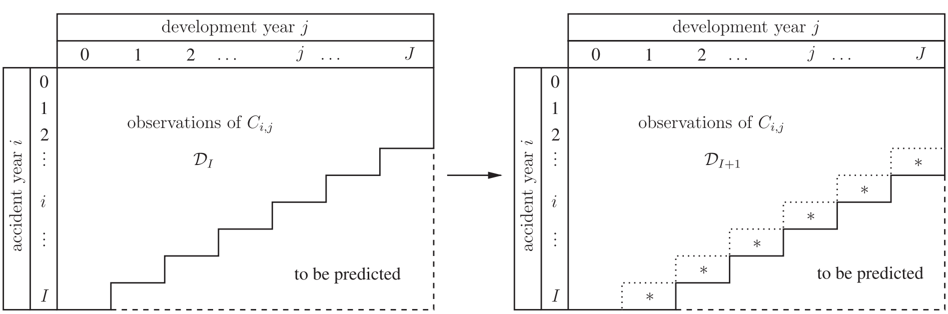

Figure 1: Information update $\mathcal{D}_I \to \mathcal{D}_{I+1}$ during the first accounting year (0,1], a new diagonal is added to the trapezoid at the end of accounting year 1.

For a further description we refer to Section 3 in Bühlmann et al. [1]. Making use of the martingale property of successive ultimate claim predictions (under the assumption that the first moments exist) we obtain from (1.1)

$$\mathbb{E}\left[\mathrm{CDR}_i \middle| \mathcal{D}_I\right] = 0.$$

Therefore, in the profit and loss statement the CDR is predicted by zero. The prediction variance (conditional mean square error of prediction) is given by, see Wüthrich-Merz [16],

$$\mathrm{msep}_{\mathrm{CDR}_i|\mathcal{D}_I}(0) = \mathbb{E}\left[\left(\mathrm{CDR}_i - 0\right)^2 \middle| \mathcal{D}_I\right] = \mathrm{Var}\left(\mathrm{CDR}_i | \mathcal{D}_I\right).$$

The prediction variance is probably one of the easiest risk measures that can be calculated. An analytical formula for this second moment is derived in Bühlmann et al. [1]. For gaining more insight into the distribution of the CDR we aim to calculate higher moments of this CDR.

2. Bayes Chain Ladder Models

In this section we consider two different distributional models for the Bayes CL model. Both models are such that the distributions belong to the exponential dispersion family with associate conjugate priors (see e.g. Bühlmann-Gisler [2]). This choice allows for closed form solutions for posterior distributions. To begin, we define the individual claims development factors by

$$F_{i,j} = \frac{C_{i,j+1}}{C_{i,j}},$$

for $i \in \{0, \ldots, I\}$ and $j \in \{0, \ldots, J-1\}$. The cumulative claims $C_{i,j}$ are then given by

$$C_{i,j+1} = C_{i,0} \prod_{k=0}^{j} F_{i,k},$$

where $C_{i,0}$ refers to the initial value of the stochastic process $(C_{i,j})_{j=0,\ldots,J}$ and the $F_{i,j}$'s describe the individual multiplicative changes from development year to development year. Next, we define the two different distributional models for the individual claims development factors $F_{i,j}$. First, we study the Gamma-Gamma Bayes CL model also considered in Salzmann-Wüthrich [15].

2.1. Gamma-Gamma Model

Model 2.1. (Gamma-Gamma Model)

Assume that $\psi_j > 0$, $j = 0, \ldots, J-1$ are given fixed constants.

(a) Conditionally, given $\boldsymbol{\Theta} = (\theta_j)_{0 \leq j \leq J-1}$, $(C_{i,j})_{j \geq 0}$ for $i = 0, \ldots, I$ are independent Markov processes with conditional distributions

$$F_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}\bigg|_{\boldsymbol{\Theta}, C_{i,j}} \sim \Gamma\left(\psi_j^{-2}, \theta_j \, \psi_j^{-2}\right).$$

(b) θ_j are independent $\Gamma(\gamma_j, f_j(\gamma_j - 1))$-distributed with given prior constants $f_j > 0$, $\gamma_j > 1$.

(c) $\boldsymbol{\Theta}$ and $C_{i,0}$ are independent and $C_{i,0} > 0$, $\mathbb{P}$-a.s.

□

Remark. The gamma distribution is denoted by $\Gamma(\gamma, c)$. In particular for $X \sim \Gamma(\gamma, c)$ with $\gamma, c > 0$ we have that $\mathbb{E}[X] = \gamma/c$ and $\operatorname{Var}(X) = \gamma/c^2$. The gamma distribution has the following moments:

$$\mathbb{E}[X^k] = c^{-k} \, \frac{\Gamma(\gamma + k)}{\Gamma(\gamma)}, \tag{2.1}$$

for $k \in \mathbb{Z}$ with $-k < \gamma$ and $\Gamma(x)$ denotes the gamma function. This determines for which k the moments of X exist.

Under Model 2.1 we have the following properties:

$$\mathbb{E}\left[C_{i,j+1} \middle| \boldsymbol{\Theta}, C_{i,j}\right] = C_{i,j} \, \mathbb{E}\left[F_{i,j} \middle| \boldsymbol{\Theta}, C_{i,j}\right] = C_{i,j} \, \theta_j^{-1}, \tag{2.2}$$

$$\operatorname{Var}\left(C_{i,j+1} \middle| \boldsymbol{\Theta}, C_{i,j}\right) = C_{i,j}^2 \operatorname{Var}\left(F_{i,j} \middle| \boldsymbol{\Theta}, C_{i,j}\right) = C_{i,j}^2 \, \psi_j^2 \, \theta_j^{-2}. \tag{2.3}$$

Remarks.

- From formula (2.2) we see that θ_j^{-1} plays the role of the development factor (see Mack [12]). The variance (2.3) is quadratic in the observation an therefore our assumptions differ from the classical distribution-free CL model of Mack [12]. As a consequence there is no prior difference in risk w.r.t. the volume between two different portfolios having the same prior parameters ψ_j and θ_j. Furthermore, the conditional coefficient of variation is given by

$$\operatorname{Vco}\left(C_{i,j+1} \,|\, \boldsymbol{\Theta}, C_{i,j}\right) = \frac{\left[\operatorname{Var}\left(C_{i,j+1} \middle| \boldsymbol{\Theta}, C_{i,j}\right)\right]^{1/2}}{\mathbb{E}\left[C_{i,j+1} \middle| \boldsymbol{\Theta}, C_{i,j}\right]} = \psi_j.$$

- The assumptions on the distributions of the θ_j's reflect our prior knowledge about the true parameters. For $\gamma_j \to 1$ we are in the non-informative case where we have no prior knowledge. Note that we need $\gamma_j > 2$ for the prior second moment of θ_j^{-1} to exist.

2.2. Lognormal-Normal Model

Secondly, we consider a lognormal distribution for the individual development factors $F_{i,j}$ and for the prior distribution we choose a normal distribution.

Model 2.2. (Lognormal-Normal Model)

Assume that $\sigma_j^2 > 0$, $j = 0, \ldots, J-1$ are given fixed constants.

(a) Conditionally, given $\boldsymbol{\mu} = (\mu_j)_{0 \le j \le J-1}$, $(C_{i,j})_{j \ge 0}$ for $i = 0, \ldots, I$ are independent Markov processes with conditional distributions

$$F_{i,j} = \frac{C_{i,j+1}}{C_{i,j}} \bigg|_{\boldsymbol{\mu}, C_{i,j}} \sim \mathcal{LN}(\mu_j, \sigma_j^2).$$

(b) μ_j are independent $\mathcal{N}(\xi_j, \nu_j^2)$-distributed with given prior constants ξ_j and $\nu_j > 0$.

(c) $\boldsymbol{\mu}$ and $C_{i,0}$ are independent and $C_{i,0} > 0$, $\mathbb{P}$-a.s.

Remarks.

- $\mathcal{LN}(\mu, \sigma^2)$ denotes the lognormal distribution and $\mathcal{N}(\mu, \sigma^2)$ the normal distribution.
- Model 2.2 inherits similar features as Model 2.1 since there is no volume term $C_{i,j}$ in the distribution of the $F_{i,j}$. The only modification is that we have different distributional assumptions.
- In Model 2.1, the gamma distributed parameter θ_j enters the model reciprocally (c.f. (2.2)-(2.3)). We know that inverse gamma distributions lie in the Fréchet maximum domain of attraction (see McNeil et al. [14]). Therefore they belong to the family of heavy-tailed distributions with infinite higher moments. As a consequence, this may lead to heavy-tailed distributions for the CDR w.r.t. $\mathcal{D}_I$ for Model 2.1, i.e. higher order moments thereof may not exist. On the other hand, as we will see in Section 3 all moments exist for the distribution of the CDR w.r.t. $\mathcal{D}_I$ resulting from the lognormal distribution of Model 2.2. We refer to such distributions as being moderately heavy-tailed, see also McNeil et al. [14].
- The structural parameters ψ_j and σ_j are assumed to be fixed constants, see Model 2.1 and Model 2.2. Only this assumption will allow for closed form solutions for higher moments of the CDR. We will derive and discuss plug-in estimators for these structural parameters in Section 4.
- Similar to Model 2.1, letting $\nu_j \to \infty$ corresponds to the non-informative case, i.e. we have no prior knowledge about the parameter μ_j.

Under Model 2.2 we have the following properties:

$$\mathbb{E}\left[C_{i,j+1} | \boldsymbol{\mu}, C_{i,j}\right] = C_{i,j}\, \mathbb{E}\left[F_{i,j} | \boldsymbol{\mu}, C_{i,j}\right] = C_{i,j} \exp\left(\mu_j + \sigma_j^2/2\right), \tag{2.4}$$

$$\mathrm{Var}\Big(C_{i,j+1}|\boldsymbol{\mu},C_{i,j}\Big) = C_{i,j}^2 \exp\Big(2\mu_j + \sigma_j^2\Big)\Big(\exp(\sigma_j^2) - 1\Big), \tag{2.5}$$

$$\mathrm{Vco}\Big(C_{i,j+1}\,|\,\boldsymbol{\mu},C_{i,j}\Big) = \Big(\exp(\sigma_j^2) - 1\Big)^{1/2}. \tag{2.6}$$

Unlike θ_j in Model 2.1, μ_j is not directly associated with the CL factor which is now given by $\exp(\mu_j + \sigma_j^2/2)$, (c.f. (2.4)) and thus depends on σ_j.

3. Posterior Distributions and Higher Moments

In this section we are going to exploit the recursive structure (see Bühlmann et al. [1]) of the different Bayes CL models presented in Section 2 in order to find closed form solutions for higher moments of the CDR, first for single accident years and second for the aggregated case.

3.1. Posterior Distributions

Since Model 2.1 and Model 2.2 belong to the exponential dispersion family with associate conjugate priors, the Bayesian estimators coincide with the linear credibility estimators (exact credibility case). Moreover, it is possible to explicitly calculate the posterior distributions of $\boldsymbol{\Theta}$ and $\boldsymbol{\mu}$, given $\mathcal{D}_{I+t}$. In the same spirit of Theorem 3.2 of Gisler-Wüthrich [6], we state the following proposition for the two different models. The proofs are provided in the appendix.

Proposition 3.1 (Exact Credibility Case). *(a) Under the assumptions of Model 2.1, given $\mathcal{D}_{I+t}$ with $t = 0, 1$, the posterior distributions of $\boldsymbol{\Theta} = (\theta_j)_{0 \le j \le J-1}$ are independent gamma distributions with parameters*

$$\gamma_{j,t} = \gamma_j + \psi_j^{-2}(I - j + t),$$

$$c_{j,t} = f_j(\gamma_j - 1) + \psi_j^{-2} \sum_{i=0}^{I-j-1+t} C_{i,j+1}.$$

(b) Under the assumptions of Model 2.2, given $\mathcal{D}_{I+t}$ with $t = 0, 1$, the posterior distributions of $\boldsymbol{\mu} = (\mu_j)_{0 \le j \le J-1}$ are independent normal distributions with parameters

$$\xi_{j,t} = \frac{\xi_j + v_j^2\sigma_j^{-2}\sum_{i=0}^{I-j-1+t}\log(F_{i,j})}{1 + (I - j + t)v_j^2\sigma_j^{-2}},$$

$$v_{j,t}^2 = \frac{v_j^2}{1 + (I - j + t)v_j^2\sigma_j^{-2}}.$$

3.2. Credibility CL Factors

This section revisits important results needed for the calculation of higher moments of the CDR.

3.2.1. *Gamma-Gamma Model*

For Model 2.1, Proposition 3.1 implies for $i > I - J + t$ that

$$\widehat{C}_{i,J}^{(t)} = \mathbb{E}[C_{i,j} | \mathcal{D}_{I+t}] = C_{i,I-i+t} \prod_{j=I-i+t}^{J-1} \mathbb{E}\left[\theta_j^{-1} \middle| \mathcal{D}_{I+t}\right] = C_{i,I-i+t} \prod_{j=I-i+t}^{J-1} \frac{c_{j,t}}{\gamma_{j,t} - 1}. \tag{3.1}$$

Define the posterior estimates of the CL factors for $j = 0, \ldots, J - 1$ by $\widehat{F}_j^{(t)} = \mathbb{E}[\theta_j^{-1} | \mathcal{D}_{I+t}]$. From straightforward calculation we obtain the following corollary:

Corollary 3.2. *Assume Model 2.1, then for* $t \le j \le J - 1$ *with* $t = 0,1$ *the posterior estimates* $\widehat{F}_j^{(t)}$ *are given by the weighted average*

$$\widehat{F}_j^{(t)} = \alpha_{j,t} \hat{f}_{j,t} + (1 - \alpha_{j,t}) f_j, \tag{3.2}$$

for $t = 0,1$ *with*

$$\hat{f}_{j,t} = \frac{\sum_{i=0}^{I-j-1+t} F_{i,j}}{I - j + t},$$

and credibility weight at time $I + t$

$$\alpha_{j,t} = \frac{I - j + t}{I - j + t + \psi_j^2 (\gamma_j - 1)}. \tag{3.3}$$

Remark. The estimator $\hat{f}_{j,t}$ corresponds to the maximum likelihood estimator (MLE) of θ_j^{-1} at time $I + t$.

Similar to Theorem 2.1 in Bühlmann et al. [1] we obtain for the posterior estimates of the CL factors in the Gamma-Gamma model the recursive structure for the updating procedure.

Corollary 3.3. *Assume Model 2.1, then for* $1 \le j \le J - 1$

$$\widehat{F}_j^{(1)} = \beta_j F_{I-j,j} + \left(1 - \beta_j\right) \widehat{F}_j^{(0)},$$

with $\mathcal{D}_I$*-measurable credibility weight*

$$\beta_j = \frac{1}{I - j + 1 + \psi_j^2 (\gamma_j - 1)}. \tag{3.4}$$

3.2.2. *Lognormal-Normal Model*

Analogous results hold true for Model 2.2, if we define $Y_{i,j} = \log F_{i,j} = \log(C_{i,j+1}/C_{i,j})$ and $\widehat{Y}_j^{(t)} = \mathbb{E}[\mu_j | \mathcal{D}_{I+t}]$. Given $\boldsymbol{\mu}$ and $C_{i,j}$, we see that $Y_{i,j}$ is normally distributed with $Y_{i,j}|_{\boldsymbol{\mu},C_{i,j}} \sim \mathcal{N}(\mu_j, \sigma_j)$ which immediately implies the next corollary.

Corollary 3.4. *Assume Model 2.1, then for* $t \leq j \leq J-1$ *with* $t = 0,1$ *the posterior estimates* $\widehat{Y}_j^{(t)}$ *are given by the weighted average*

$$\widehat{Y}_j^{(t)} = \xi_{j,t} = \rho_{j,t} \bar{Y}_j^{(t)} + (1 - \rho_{j,t}) \xi_j, \tag{3.5}$$

with

$$\begin{aligned} \bar{Y}_j^{(t)} &= \frac{1}{I-j+t} \sum_{i=0}^{I-j-1+t} Y_{i,j} \qquad \textit{(MLE of } \mu_j \textit{ at time } I+t\textit{)}, \\ \rho_{j,t} &= \frac{(I-j+t) v_j^2 \sigma_j^{-2}}{1 + (I-j+t) v_j^2 \sigma_j^{-2}} \qquad \textit{(credibility weight at time } I+t\textit{)}. \end{aligned} \tag{3.6}$$

Moreover, we obtain the following recursive structure in this case.

Corollary 3.5. *Assume Model 2.2, then for* $1 \leq j \leq J-1$

$$\widehat{Y}_j^{(1)} = \kappa_j Y_{I-j,j} + (1-\kappa_j) \widehat{Y}_j^{(0)},$$

with credibility weight

$$\kappa_j = \frac{v_j^2 \sigma_j^{-2}}{1 + (I-j+1) v_j^2 \sigma_j^{-2}}. \tag{3.7}$$

Due to the assumptions of Model 2.2 and the posterior independence of the parameter $\boldsymbol{\mu} = (\mu_j)_{0 \leq j \leq J-1}$ according to Proposition 3.1 we thus obtain

$$\widehat{C}_{i,J}^{(t)} = \mathbb{E}\left[C_{i,J} \middle| \mathcal{D}_{I+t}\right] = C_{i,I-i+t} \prod_{j=I-i+t}^{J-1} \mathbb{E}\left[\exp\left(\mu_j + \frac{\sigma_j^2}{2}\right) \middle| \mathcal{D}_{I+t}\right] \tag{3.8}$$

$$= C_{i,I-i+t} \prod_{j=I-i+t}^{J-1} \exp\left(\xi_{j,t} + \frac{v_{j,t}^2}{2}\right) \exp\left(\frac{\sigma_j^2}{2}\right). \tag{3.9}$$

Moreover, with Corollary 3.5 we get

$$\widehat{F}_j^{(1)} = \mathbb{E}\left[\exp\left(\mu_j + \frac{\sigma_j^2}{2}\right) \middle| \mathcal{D}_{I+1}\right] = \exp\left(\kappa_j Y_{I-j,j} + (1-\kappa_j) \widehat{Y}_j^{(0)}\right) \exp\left(\frac{\sigma_j^2}{2} + \frac{v_{j,1}^2}{2}\right). \tag{3.10}$$

Note that in the above formula $Y_{I-j,j}$ is the only term which is not $\mathcal{D}_I$-measurable.

3.3. Higher Moments Single Accident Years

Proposition 3.1, Corollary 3.3 and Corollary 3.5 allow for the calculation of higher moments of the ultimate claim predictor $\widehat{C}_{i,J}^{(1)}$ given $\mathcal{D}_I$ for both models.

Proposition 3.6. *Choose $i > I - J$.*

(a) *Assume Model 2.1 and let $N_i = \min\{\gamma_{I-i,0}, \dots, \gamma_{J-1,0}\}$. For $n \in \mathbb{N}$ with $n < N_i$ we obtain*

$$\mathbb{E}\left[\left(\widehat{C}_{i,J}^{(1)}\right)^n \middle| \mathcal{D}_I\right] = \left(C_{i,I-i}\right)^n K_{n,I-i} \prod_{j=I-i+1}^{J-1} \left[\sum_{l=0}^{n} \binom{n}{l} (\beta_j)^l K_{l,j} (1-\beta_j)^{n-l} \left(\widehat{F}_j^{(0)}\right)^{n-l}\right],$$

where credibility weight β_j is given by (3.4) *and for $j = I - i, \dots, J - 1$ and $0 \le l \le n$*

$$K_{l,j} = \left(\psi_j^2 c_{j,0}\right)^l \frac{\Gamma\left(\psi_j^{-2} + l\right)}{\Gamma\left(\psi_j^{-2}\right)} \frac{\Gamma(\gamma_{j,0} - l)}{\Gamma(\gamma_{j,0})},$$

with $c_{j,0}$ and *$\gamma_{j,0}$ given by Proposition 3.1.*

(b) Assume Model 2.2. We obtain

$$\mathbb{E}\left[\left(\widehat{C}_{i,J}^{(1)}\right)^n \middle| \mathcal{D}_I\right] = C_{i,I-i}^n \prod_{j=I-i}^{J-1} \exp(n\xi_{j,0}) M_{n,j},$$

with

$$M_{n,j} = \exp\left(\frac{(n\kappa_j)^2 \sigma_j^2}{2} + \frac{(n\kappa_j)^2 v_{j,0}^2}{2} + n\left(\frac{\sigma_j^2}{2} + \frac{v_{j,1}^2}{2}\right) \mathbb{1}_{\{j > I-i\}}\right),$$

with $\xi_{j,0}$ and $v_{j,t}$ given by Proposition 3.1 and κ_j given by (3.7) *for $j > I - i$ and $\kappa_{I-i} = 1$.*

Remarks.

- In Proposition 3.6 we calculate moments of $\widehat{C}_{i,J}^{(1)}$ given $\mathcal{D}_I$. To get moments of CDR_i we have to consider moments of $\widehat{C}_{i,J}^{(0)} - \widehat{C}_{i,J}^{(1)}$. Since $\widehat{C}_{i,J}^{(0)}$ is $\mathcal{D}_I$-measurable this can be done easily. For instance, the second moment of CDR_i is given by

$$\mathbb{E}[\mathrm{CDR}_i^2 | \mathcal{D}_I] = \mathbb{E}\left[\left(\widehat{C}_{i,J}^{(0)} - \widehat{C}_{i,J}^{(1)}\right)^2 \middle| \mathcal{D}_I\right] = \mathbb{E}\left[\left(\widehat{C}_{i,J}^{(1)}\right)^2 \middle| \mathcal{D}_I\right] - \left(\widehat{C}_{i,J}^{(0)}\right)^2.$$

Therefore, moments of CDR_i given $\mathcal{D}_I$ are determined by moments of $\widehat{C}_{i,J}^{(1)}$.

- Moreover we can calculate the skewness and excess kurtosis of CDR_i. In the case study in Section 5 we change signs for the CDR such that losses in the CDR correspond to positive values (and are in-line with the classical definition of risk measures). Therefore, the skewness and excess kurtosis thereof are given by

$$\text{skewness} = \frac{\mathbb{E}[-\mathrm{CDR}_i^3|\mathcal{D}_I]}{(\mathbb{E}[\mathrm{CDR}_i^2|\mathcal{D}_I])^{3/2}}, \quad \text{excess kurtosis} = \frac{\mathbb{E}[\mathrm{CDR}_i^4|\mathcal{D}_I]}{(\mathbb{E}[\mathrm{CDR}_i^2|\mathcal{D}_I])^2} - 3.$$

  Note that the moments in the above formulas for the skewness and the excess kurtosis are moments relative to the mean which in our case is given by $\mathbb{E}[\mathrm{CDR}_i|\mathcal{D}_I] = 0$.

- Skewness smaller than zero may indicate distributions with heavier left tail than right tail whereas skewness greater than zero may indicate that the right tail is heavier. Excess kurtosis measures the excess of the kurtosis over the kurtosis of a normal distribution. Excess kurtosis greater than zero indicates that the distribution is more peaked and has a heavier tail than a normal distribution.

Linear combinations of independent normally distributed random variables are again normally distributed. It is therefore possible for Model 2.2 to get a closed form solution for the distribution of $\widehat{C}_{i,J}^{(1)}$, given $\boldsymbol{\mu}$ and $\mathcal{D}_I$, according to the following proposition.

Proposition 3.7. *Choose $i > I - J$. Assume Model 2.2. Conditionally, given $\boldsymbol{\mu}$ and $\mathcal{D}_I$, $\widehat{C}_{i,J}^{(1)}$ has the following distribution*

$$\widehat{C}_{i,J}^{(1)}|_{\boldsymbol{\mu},\mathcal{D}_I} \sim \mathcal{LN}\left(d_i + \sum_{j=I-i}^{J-1} \kappa_j \mu_j, \sum_{j=I-i}^{J-1} \kappa_j^2 \sigma_j^2\right),$$

with κ_j given by (3.7) *for $j > I - i$ and $\kappa_{I-i} = 1$ and*

$$d_i = \log(C_{i,I-i}) + \mathbb{1}_{\{i > I-J+1\}} \sum_{j=I-i+1}^{J-1} \left(\kappa_j \sum_{k=0}^{I-j-1} \log(F_{k,j}) + \frac{\xi_j}{1 + (I-j+1)\nu_j^2 \sigma_j^{-2}} + \frac{\sigma_j^2}{2} + \frac{\nu_{j,1}^2}{2}\right).$$

3.4. Higher Moments Aggregated Accident Years

To study uncertainties in the CDR for aggregated accident years, we consider moments of $\widehat{C}_{\bullet,J}^{(1)}$ defined by

$$\widehat{C}_{\bullet,J}^{(1)} = \sum_{i=I-J+1}^{I} \widehat{C}_{i,J}^{(1)}.$$

The following theorem gives closed form solutions for both models.

Theorem 3.8. *For $n \in \mathbb{N}$ and $k_0, \ldots, k_{J-1} \in \{0, 1, \ldots, n\}$ define $h_s = \sum_{u=0}^{s} k_u$*

(a) Assume Model 2.1. For $n < \min\{\gamma_{0,0}, \ldots, \gamma_{J-1,0}\}$ we obtain

$$\mathbb{E}\left[\left(\widehat{C}^{(1)}_{\bullet,J}\right)^n \middle| \mathcal{D}_I\right] = \sum_{\substack{k_0,\ldots,k_{J-1}\in\{0,\ldots,n\}\\ h_{J-1}=n}} \binom{n}{k_0, k_1, \ldots, k_{J-1}} \prod_{s=0}^{J-1} \Bigg(C^{k_s}_{I-J+1+s,\, J-1-s}$$
$$\times \sum_{r=0}^{n-h_s} \binom{n-h_s}{r} \beta^r_{J-1-s} \left((1-\beta_{J-1-s})\, \widehat{F}^{(0)}_{J-1-s}\right)^{n-h_s-r} K_{k_s+r,\, J-1-s}\Bigg),$$

where $K_{l,j}$ are given by Proposition 3.6 (a), β_j is given by (3.4) *and the last sum is set equal to $K_{k_s,\, J-1-s}$ if $n - h_s = 0$.*

(b) Assume Model 2.2. We obtain

$$\mathbb{E}\left[\left(\widehat{C}^{(1)}_{\bullet,J}\right)^n \middle| \mathcal{D}_I\right] = \sum_{\substack{k_0,\ldots,k_{J-1}\in\{0,\ldots,n\}\\ h_{J-1}=n}} \binom{n}{k_0, k_1, \ldots, k_{J-1}} \prod_{s=0}^{J-1} \Bigg(C^{k_s}_{I-J+1+s,\, J-1-s}$$
$$\times \exp\left((k_s + (n-h_s))\, \xi_{J-1-s,0}\right) H_s\Bigg),$$

where

$$H_s = \exp\Bigg(\frac{\left(k_s + \kappa_{J-1-s}(n-h_s)\right)^2}{2} \left(\sigma^2_{J-1-s} + v^2_{J-1-s,0}\right) + \frac{(n-h_s)}{2} \left(\sigma^2_{J-1-s} + v^2_{J-1-s,1}\right)\Bigg),$$

and κ_j is given by (3.7).

4. Estimation of Structural Parameters

The structural parameters ψ_j and σ_j play an important role in the two models and have strong implications on the distribution of the individual claims development factors $F_{i,j}$. Furthermore, if one has no prior knowledge about these constants (e.g. expert opinion or industry wide data), the most reasonable thing to do is to estimate them from internal data. To model them stochastically similar to θ_j or μ_j is no feasible alternative as we will explain later. Therefore, we need to make some remarks about the estimation of ψ_j and σ_j.

4.1. Maximum a Posteriori Estimators

Since there is no canonical method how to estimate the constants ψ_j and σ_j, we rely on a semi ad-hoc method based on maximum a posteriori (MAP) estimators which are canonical estimators within a Bayesian framework. We start with Model 2.2 since this setup allows for an analytical solution. Similarly, we can determine estimators for Model 2.1 but for their calculation we have to rely on numerical optimization techniques.

Remark. The assumptions of Model 2.1 and Model 2.2 allow for a mathematically consistent methodology to determine posterior estimates of θ_j^{-1} and μ_j (i.e. $\widehat{F}_j^{(t)}$ and $\widehat{Y}_j^{(t)}$, c.f. (3.2) and (3.5)), respectively. These minimum mean squared error (MMSE) estimators for θ_j^{-1} and μ_j involve the structural parameters ψ_j and σ_j, respectively, which are assumed to be given fixed constants in Model 2.1 and Model 2.2. This means that we have to determine the constants ψ_j and σ_j before we can calculate MMSE estimators for θ_j^{-1} and μ_j. As we will see MAP estimators for ψ_j and σ_j depend on θ_j^{-1} and μ_j, respectively, which leads to implicit solutions only. Therefore we have to calculate MAP estimators for θ_j^{-1} and μ_j as well to get MAP estimators for ψ_j and σ_j which we then use as plug-in estimators for ψ_j and σ_j in order to obtain MMSE estimators for θ_j^{-1} and μ_j in Model 2.1 and Model 2.2.

Model 2.2: To apply the MAP method, we have to choose a prior distribution for σ_j^2. Having no prior information about the parameter we simply assume that all possible values for σ_j^2 lie in an interval $[0, M]$ and are equally likely. Therefore, we assume that σ_j are independent uniformly distributed on $[0, M]$. Hence under Model 2.2 we obtain

$$\begin{aligned} L_{\mathcal{D}_I}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2) \propto & \prod_{i+j\leq I-1} \frac{1}{F_{i,j}\sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{1}{2}\frac{(\log(F_{i,j})-\mu_j)^2}{\sigma_j^2}\right\} \\ & \times \prod_{j=0}^{J-1} \frac{1}{\sqrt{2\pi\nu_j^2}} \exp\left\{-\frac{1}{2}\left(\frac{\mu_j-\xi_j}{\nu_j}\right)^2\right\} \frac{1}{M}\mathbb{1}_{[0,M]}(\sigma_j^2). \end{aligned}$$

In the case study of Section 5 we analyze a portfolio of one line of business (LOB) which is split into different business units (BU), e.g. into different regions. Since we assume that BU's have the same prior underlying risk characteristics, we will work with overall portfolio estimators for the LOB. This means that we assume that each BU conforms to the same model with the same prior parameters and the same fixed constants. Denote by n the number of BU's and assume that these are independent conditional on the model parameters. This provides that the joint posterior distributions of $(\boldsymbol{\mu}, \boldsymbol{\sigma}^2) = (\mu_j, \sigma_j^2)_{0\leq j\leq J-1}$ are independent. Therefore, we consider the density

$$\begin{aligned} L_{\mathcal{D}_I}^{LOB}(\mu_j, \sigma_j^2) \propto & \prod_{k=1}^{n} \prod_{i\leq I-j-1} \frac{1}{F_{i,j}^{(k)}\sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{1}{2}\frac{\left(\log(F_{i,j}^{(k)})-\mu_j\right)^2}{\sigma_j^2}\right\} \\ & \times \frac{1}{\sqrt{2\pi\nu_j^2}} \exp\left\{-\frac{1}{2}\left(\frac{\mu_j-\xi_j}{\nu_j}\right)^2\right\} \frac{1}{M}\mathbb{1}_{[0,M]}(\sigma_j^2), \end{aligned} \tag{4.1}$$

for the estimation of the overall parameters μ_j and σ_j^2. $F_{i,j}^{(k)}$ denotes the individual development factor of the k-th BU for $k \in \{1, \ldots, n\}$. To simplify calculation we assume that M is chosen sufficiently large, such that the mode in σ_j^2 lies in $[0, M]$, and we set the indicator function in formula (4.1) equal to 1. The MAP parameters are then estimated by finding μ_j and σ_j^2 such that

$$\left(\widehat{\mu_j}^{MAP}, \widehat{\sigma_j^2}^{MAP}\right) = \underset{\mu_j, \sigma_j^2}{\operatorname{argmax}} \, L_{\mathcal{D}_I}^{LOB}(\mu_j, \sigma_j^2).$$

By considering the log-likelihood function w.r.t. $\mathcal{D}_I$ and setting the first partial derivatives equal to zero we have to solve the following system of equations

$$\widehat{\mu_j}^{MAP} = \frac{1}{n(I-j) + \widehat{\sigma_j^2}^{MAP} v_j^{-2}} \left(\sum_{k=1}^{n} \sum_{i \le I-j-1} \log(F_{i,j}^{(k)}) + \widehat{\sigma_j^2}^{MAP} v_j^{-2} \xi_j \right), \quad (4.2)$$

$$\widehat{\sigma_j^2}^{MAP} = \frac{1}{n(I-j)} \sum_{k=1}^{n} \sum_{i \le I-j-1} \left(\log(F_{i,j}^{(k)}) - \widehat{\mu_j}^{MAP} \right)^2. \quad (4.3)$$

The solution thereof involves solving a polynomial equation of order 3. However, for non-informative priors, i.e. $v_j^2 \to \infty$, the terms including v_j^{-2} in (4.2) vanish and the MAP estimator for μ_j converges in limit to the MLE for μ_j. Therefore we obtain in the non-informative prior case

$$\widehat{\mu_j}^{MAP} \approx \widehat{\mu_j}^{MLE} = \frac{1}{n(I-j)} \sum_{k=1}^{n} \sum_{i \le I-j-1} \log(F_{i,j}^{(k)}). \quad (4.4)$$

By plugging in (4.4) in (4.3) we get an estimator for σ_j^2 in the case of non-informative priors.

Model 2.1: For Model 2.1 we consider the following posterior likelihood function over all BU's, i.e. $k \in \{1, \ldots, n\}$

$$L_{\mathcal{D}_I}^{LOB}(\theta_j, \psi_j^2) \propto \prod_{k=1}^{n} \prod_{i \le I-j-1} \frac{\left(\theta_j \psi_j^{-2}\right)^{\psi_j^{-2}}}{\Gamma\left(\psi_j^{-2}\right)} \left(F_{i,j}^{(k)}\right)^{\psi_j^{-2}-1} \exp\left\{-\theta_j \psi_j^{-2} F_{i,j}^{(k)}\right\}$$
$$\times \frac{(f_j(\gamma_j - 1))^{\gamma_j}}{\Gamma(\gamma_j)} \theta_j^{\gamma_j - 1} \exp\{-f_j(\gamma_j - 1)\theta_j\} \frac{1}{M} 1_{[0,M]}(\psi_j^2).$$

In this case we cannot find analytical solutions for $\widehat{\theta_j}^{MAP}$ and $\widehat{\psi_j^2}^{MAP}$. Therefore, with the same simplification as above, i.e. setting the indicator function equal to 1, we have to numerically solve the optimization problem

$$(\widehat{\theta_j}^{MAP}, \widehat{\psi_j^2}^{MAP}) = \underset{\theta_j, \psi_j^2}{\operatorname{argmax}} \, L_{\mathcal{D}_I}^{LOB}(\theta_j, \psi_j^2).$$

4.2. Ad-hoc Estimators

In the following section, we derive another set of ad-hoc estimators for the structural parameters. These estimators serve as reference values and we use them to initialize the optimization algorithm of the MAP estimators for Model 2.1. In the case study of Section 5 we are going to compare results for the MAP estimators with results for the ad-hoc estimators.

Model 2.1: First we derive the estimators for single BU's. From the coefficient of variation we obtain for each BU for $k \in \{1, \ldots, n\}$

$$\psi_j^{2(k)} = \frac{\mathrm{Var}\left[F_{i,j}^{(k)} \middle| \boldsymbol{\Theta}, \mathcal{D}_I\right]}{\mathbb{E}\left[F_{i,j}^{(k)} \middle| \boldsymbol{\Theta}, \mathcal{D}_I\right]^2}.$$

From this we derive estimators for $\psi_j^{2(k)}$ by

$$\widehat{\psi_j^2}^{(k)} = \frac{\frac{1}{I-j-1}\sum_{i=0}^{I-j-1}\left(F_{i,j}^{(k)} - \hat{f}_{j,0}^{(k)}\right)^2}{\left(\hat{f}_{j,0}^{(k)}\right)^2}, \tag{4.5}$$

where $\hat{f}_{j,0}^{(k)}$ is given by Corollary 3.2. To get overall (LOB) estimators for ψ_j^2, we take the arithmetic mean of the estimators (4.5) over all single BU's (see Bühlmann-Gisler [2], Section 4.8), i.e.

$$\widehat{\psi_j^2}^{\text{ad-hoc}} = \frac{1}{n}\sum_{k=1}^{n} \widehat{\psi_j^2}^{(k)}.$$

Model 2.2: Similar to Model 2.1, Model 2.2 has variance proportional to the mean squared (c.f. (2.4) and (2.5)). Therefore the CL factor is estimated by the empirical mean as presented in Mack [13]. Hence we introduce

$$\bar{F}_j^{(k)} = \frac{1}{I-j}\sum_{i=0}^{I-j-1} F_{i,j}^{(k)},$$

and derive an estimator for the coefficient of variation (2.6) by

$$\widehat{\left(\exp(\sigma_j^2) - 1\right)}^{(k)} = \frac{\frac{1}{I-j-1}\sum_{i=0}^{I-j-1}\left(F_{i,j}^{(k)} - \bar{F}_j^{(k)}\right)^2}{\left(\bar{F}_j^{(k)}\right)^2}.$$

From this we derive an estimator for σ_j^2 w.r.t. the k-th BU by

$$\widehat{\sigma_j^2}^{(k)} = \log\left(\frac{\frac{1}{I-j-1}\sum_{i=0}^{I-j-1}\left(F_{i,j}^{(k)} - \bar{F}_j^{(k)}\right)^2}{\left(\bar{F}_j^{(k)}\right)^2} + 1\right).$$

Again, we take the arithmetic mean over all BU's to get the overall estimator $\widehat{\sigma_j^2}^{\text{ad-hoc}}$.

Remark. By using the estimators above we lose nice properties, e.g. unbiasedness. However, since there is no feasible alternative we have to rely on such methods. To mitigate this weakness one could think of considering full Bayes models, i.e. all the necessary parameters are modeled stochastically. As a consequence, it is then no longer possible to calculate the posterior distributions in closed form as in Proposition 3.1, i.e. one has to proceed with Markov Chain Monte Carlo methods. Because of nested structures direct simulation of the CDR turns out to be too time consuming. Therefore, we cannot exploit full Bayes models for an investigation of the distribution of the CDR, unless we have a closed form solution for the CDR.

5. Case Study

In the following case study we analyze the data-set from Gisler-Wüthrich [6], pages 598-600 which records one LOB that is separated into 6 different regions, i.e. $\text{BU}_A, \ldots, \text{BU}_F$. Mainly due to different volumes and scarce data, we detect strong random fluctuations in the claims development pattern on a regional level (see Gisler-Wüthrich [6], Section 8, Numerical Example). First of all, we determine the prior parameters f_j, γ_j, ξ_j and v_j and the MAP estimates as well as the ad-hoc estimates for the squares of the structural parameters ψ_j and σ_j (see Section 4). Table 1 gives an overview of the parameters for Model 2.1 and Model 2.2. For Model 2.1 we use the same prior values for f_j as in Gisler-Wüthrich [6], which are derived from similar techniques as in e.g. Section 4.8 of Bühlmann-Gisler [2]. Furthermore, note that we work with vague priors for $\boldsymbol{\Theta}$ and $\boldsymbol{\mu}$, i.e. γ_j and v_j (see Table 1) are chosen such that the credibility weights $\alpha_{j,0}$ and $\rho_{j,0}$, respectively, (c.f. (3.3) and (3.6)) are close to 100%.

As described in Section 4, we take the arithmetic mean of the ad-hoc estimators for single BU's to get an overall ad-hoc estimator for the LOB. In particular, this means that despite BU_F shares only a relatively small proportion of the whole portfolio it contributes equally to the total uncertainty of the LOB. Since BU_F is subject to strong random fluctuations (see also Gisler-Wüthrich [6]) this partly explains the higher values for ad-hoc estimators for $j = 0, 1$ compared to MAP estimators.

5.1. Single Accident Years

First we provide an analysis of higher moments of single accident years. In particular, by considering the skewness and the excess kurtosis of CDR_i we get an idea about the shape and the riskiness of the distribution of the CDR for each single accident year. The moments of the CDR depend on the estimation/choice of the structural parameters. Therefore we compare the results for the MAP estimators with the results for the ad-hoc estimators from Section 4.

TABLE 1

PRIOR PARAMETERS AND ESTIMATES OF STRUCTURAL PARAMETERS FOR BOTH MODELS. THE CREDIBILITY WEIGHTS ARE CALCULATED W.R.T. $\widehat{\psi_j^2}^{MAP}$ AND $\widehat{\sigma_j^2}^{MAP}$, RESPECTIVELY.

j	0	1	2	3	4	5	6	7	8	9
Model 2.1										
f_j	2.111	1.129	1.033	1.013	1.004	1.001	0.993	0.998	1.000	0.999
$(\widehat{\theta_j}^{MAP})^{-1}$	3.053	1.182	1.056	1.016	0.996	1.002	0.996	1.002	1.000	1.001
γ_j	2.1	2.1	2.2	2.1	3.4	2.1	2.1	3.4	7.7	7.5
$\widehat{\psi_j^2}^{MAP}$	0.4270	0.0515	0.0119	0.0041	0.0004	0.0025	0.0024	0.0001	0.0001	0.0002
$\widehat{\psi_j^2}^{\text{ad-hoc}}$	0.6891	0.0773	0.0106	0.0049	0.0003	0.0025	0.0020	0.0001	0.0001	0.0002
$\alpha_{j,0}$	97.59%	99.69%	99.92%	99.97%	99.99%	99.98%	99.98%	100.00%	99.99%	99.99%
Model 2.2										
ξ_j	0.8900	0.1411	0.0483	0.0139	0.0030	0.0005	−0.0005	−0.0002	−0.0010	−0.0006
$\widehat{\mu_j}^{MAP}$	0.8947	0.1411	0.0483	0.0139	0.0029	0.0005	−0.0048	−0.0002	−0.0010	−0.0006
v_j	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
$\widehat{\sigma_j^2}^{MAP}$	0.2270	0.0384	0.0108	0.0035	0.0004	0.0024	0.0025	0.0001	0.0001	0.0002
$\widehat{\sigma_j^2}^{\text{ad-hoc}}$	0.3250	0.0682	0.0104	0.0048	0.0003	0.0025	0.0020	0.0001	0.0001	0.0002
$\rho_{j,0}$	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

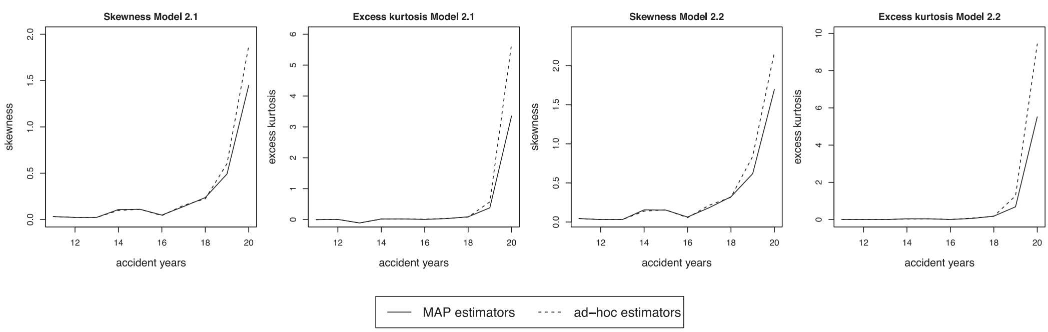

FIGURE 2: Skewness and excess kurtosis of CDR_i for accident years $i = 11, \ldots, 20$ for Model 2.1 and Model 2.2 w.r.t. MAP estimators (solid line) and w.r.t. ad-hoc estimators (dashed line).

Figure 2 summarizes the results for Model 2.1 and Model 2.2. Since for both models there is no volume term in the parametrization of the distribution of the individual development factors $F_{i,j}$ we observe almost identical results for all different BU's and therefore it is sufficient to consider the results for BU_A as a representative. Skewness and excess kurtosis are increasing for more recent accident years which is in-line with the observation that most of the development takes place in the first development years. Moreover we see that the results are sensitive to the estimation/choice of ψ_j^2 and σ_j^2. Especially for the most recent accident years $i = 19,20$, we observe that higher values of $\widehat{\psi_j^2}^{\text{ad-hoc}}$ and $\widehat{\sigma_j^2}^{\text{ad-hoc}}$, compared to $\widehat{\psi_j^2}^{MAP}$ and $\widehat{\sigma_j^2}^{MAP}$, lead to higher skewness and excess kurtosis. As mentioned in Section 2, Model 2.1 involves the inverse gamma distribution which may lead to heavy-tailed distributions for the CDR given $\mathcal{D}_I$ and higher moments thereof may fail to exist. In fact for this example we observe that the minimum order (over all accident years and BU's) up to which moments exist is $n = 48$ w.r.t. MAP estimators and $n = 31$ w.r.t. ad-hoc estimators. On the other hand we point out that for Model 2.2 all moments exist.

5.2. Aggregated Accident Years

In the following we consider the results for aggregated accident years. We calculate higher moments according to Theorem 3.8 and derive the values for the skewness and the excess kurtosis. To highlight the influence of the consideration of skewness and excess kurtosis in the uncertainty analysis of the CDR we then fit a distribution from the 4-parameter Johnson family of distributions (see Johnson et al. [11]), which includes the shifted lognormal distribution as a special case, to the first four moments of the CDR and compare the results for the standard shifted lognormal fit taking only the first two moments into account as proposed in Solvency II. The Johnson fit can either be calculated by the help of tables (see Johnson [10]) or by iterative methods as described in Elderton-Johnson [4]. For this case study we made use of the function `JohnsonFit` implemented in the `R`-package `SuppDists` for which a fit by moments, presented in Hill et al. [8], is applied. In the following we restrict our consideration to results w.r.t. MAP estimators which are more robust compared to ad-hoc estimators.

From Table 2 we observe skewness to the right and positive excess kurtosis in all cases. The distribution fit resulting from `JohnsonFit` lies in the so-called S_U-system represented by

$$Z = \gamma + \delta \sinh^{-1}\left\{\frac{X-\xi}{\lambda}\right\},$$

where Z has standard normal distribution, for more details see Johnson et al. [11]. As documented in the help file of the `R`-package `SuppDists` fitting by moments is difficult and the function `JohnsonFit` may report an error in some cases which happens for BU_A (see Table 2) in our example. For the other BU's we see that the fitted Johnson distributions fit the skewness and excess kurtosis very close whereas the values for the fitted shifted lognormal distributions significantly deviate. Note that the shifted lognormal fit currently used in insurance industry provides substantially lower skewness and excess kurtosis. As a representative for all BU's Figure 3 illustrates the fitted densities for BU_C for Model 2.1 and Model 2.2.

In summary, we calculate exact moments for Model 2.1 and Model 2.2. For these moments we then fit the standard shifted lognormal distribution to the

TABLE 2

SKEWNESS (S) AND EXCESS KURTOSIS (EK) OF THE AGGREGATED CDR W.R.T. THEOREM 3.8 FOR THE FITTED JOHNSON DISTRIBUTION ($\mathcal{J}$) AND FOR THE FITTED SHIFTED LOGNORMAL DISTRIBUTION ($\mathcal{LN}$) W.R.T. MAP ESTIMATORS.

	Model 2.1						Model 2.2					
	S	$S_{\mathcal{J}}$	$S_{\mathcal{LN}}$	EK	$EK_{\mathcal{J}}$	$EK_{\mathcal{LN}}$	S	$S_{\mathcal{J}}$	$S_{\mathcal{LN}}$	EK	$EK_{\mathcal{J}}$	$EK_{\mathcal{LN}}$
BU_A	1.36	NA	0.23	3.08	NA	0.09	1.56	1.57	0.18	4.92	4.93	0.06
BU_B	0.79	0.79	0.10	1.42	1.41	0.02	0.80	0.81	0.09	1.88	1.90	0.01
BU_C	1.01	1.01	0.13	2.01	2.00	0.03	1.12	1.11	0.11	3.03	3.01	0.02
BU_D	0.89	0.89	0.10	1.68	1.68	0.02	0.96	0.97	0.09	2.43	2.44	0.01
BU_E	0.75	0.75	0.12	1.22	1.22	0.02	0.82	0.83	0.10	1.78	1.81	0.02
BU_F	0.74	0.74	0.11	1.30	1.30	0.02	0.31	0.32	0.08	0.40	0.41	0.01

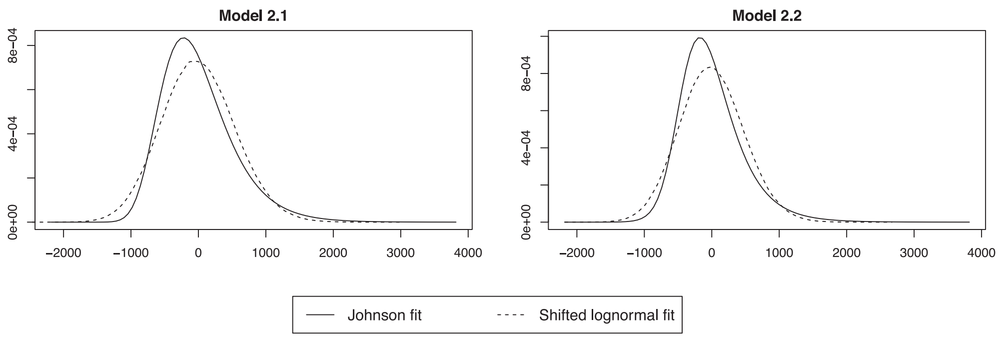

FIGURE 3: Fitted densities for the Johnson family of distributions (solid line) and the shifted lognormal distribution (dashed line) for Model 2.1 and Model 2.2 w.r.t. BU_C and MAP estimators.

first two moments as proposed in Solvency II. Comparing this approximation to the 4-parameter Johnson approximation, using the exact skewness and excess kurtosis resulting from Model 2.1 and Model 2.2 respectively, shows that the approximation for the standard shifted lognormal distribution is too symmetric and fails to fit the skewness and the excess kurtosis of the distribution of the CDR (see Table 2). This has significant influence in setting confidence levels or choosing risk adjustment estimation techniques as discussed in IASB [9]: Figure 4 shows quantiles for the Johnson and the shifted lognormal approximation as a function of the confidence level for both models w.r.t. BU_C. For both models the quantiles of the Johnson fit are smaller than the quantiles of the shifted lognormal approximation for confidence levels below 90%. Slightly above the 90% confidence level they become equal and then start to significantly deviate (values up to confidence level 99.5% used in Solvency II are provided), where the quantiles of the Johnson fit are larger compared to the quantiles of the shifted lognormal approximation.

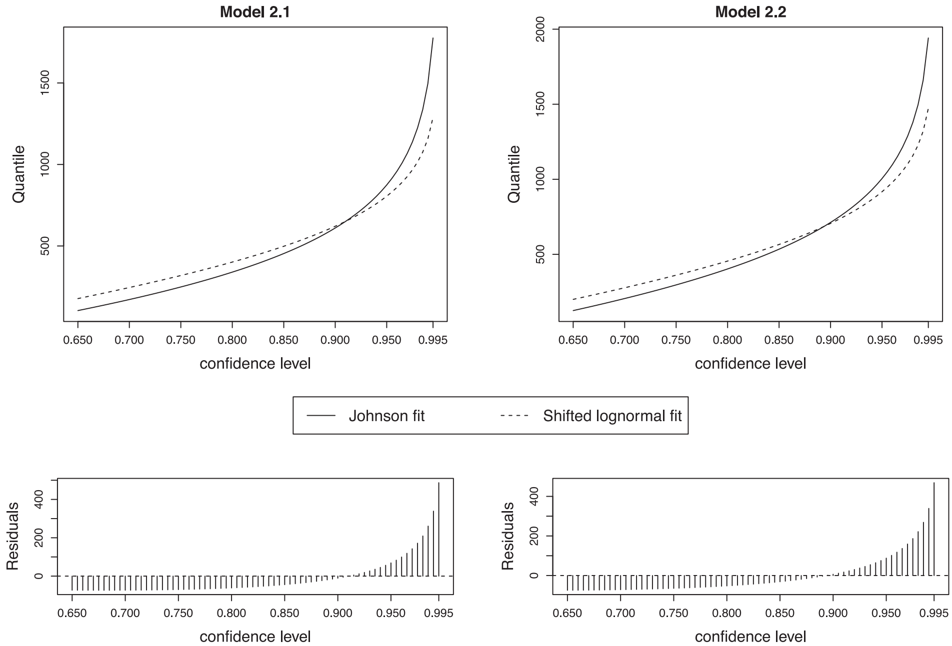

FIGURE 4: Quantiles for the Johnson fit (solid line) and quantiles for the shifted lognormal distribution (dashed line) as a function of the confidence level w.r.t. BU_C and MAP estimators. Residuals of quantiles for the two approximations.

As an example we think of two insurers (Insurer A and B) assessing the uncertainty of the CDR for BU_C. Insurer A calculates moments up to order four (e.g. calculated w.r.t. Model 2.1) and considers the Johnson fit and Insurer B calculates the first two moments (w.r.t. Model 2.1) and considers the shifted lognormal fit. If both choose a quantile slightly on the 90% confidence level as risk measure they will report about equal risk adjustments. We know that the two approximations result in different values for skewness and excess kurtosis. This means that in our example the requirement of paragraph B94 in IASB [9] that risk adjustments have to be larger for probability distributions that are more skewed would be violated. Insurer A that calculated moments up to order four knows that the distribution of the CDR is rather skewed and according to paragraph B95 in IASB [9] he concludes that the confidence level technique is not appropriate and considers other techniques, e.g. the conditional tail expectation, to estimate the risk adjustment. On the other hand Insurer B that considers the shifted lognormal fit may wrongly account the confidence level technique to be viable. Therefore, taking factors such as the shape of the distribution into account significantly influences the assessment of the uncertainty of the CDR as discussed in IASB [9].

Considering high quantiles for the two approximations, e.g. the VaR at the 99.5% confidence level listed in Table 3, shows that the VaR for the shifted lognormal distribution is significantly lower compared to the VaR for the Johnson

TABLE 3

VaR at 99.5% confidence level and reserves for Model 2.1 and Model 2.2 w.r.t. MAP estimators, for the fitted Johnson distribution ($\mathcal{J}$) and the fitted shifted lognormal distribution ($\mathcal{LN}$).

	Model 2.1			Model 2.2		
	$\text{VaR}_{\mathcal{J}}$	$\text{VaR}_{\mathcal{LN}}$	reserves	$\text{VaR}_{\mathcal{J}}$	$\text{VaR}_{\mathcal{LN}}$	reserves
BU_A	NA	2110	815	2429	1668	847
BU_B	1075	849	321	963	741	372
BU_C	1942	1473	748	1776	1289	852
BU_D	3846	2962	1498	3517	2610	1708
BU_E	2081	1674	570	1914	1479	716
BU_F	208	167	114	123	111	77

fit in all cases. However, the first four moments do not tell much about the tail behavior of the distribution of the CDR. Therefore a comprehensive assessment of the risk in the tails should take into account techniques from extreme value theory as e.g. presented in Embrechts et al. [5].

Furthermore, we see that the VaR at the 99.5% confidence level for Model 2.1 is significantly higher than for Model 2.2 for all BU's. This can partly be explained by the fact that Model 2.1 involves inverse gamma distributions leading to heavy tailed distributions for the CDR for which higher moments fail to exist whereas Model 2.2 results in moderately heavy tailed distributions for which all moments exist.

Remarks.

- Another approach to estimate the VaR taking moments up to order four into account, is provided by the Cornish-Fisher expansion, see e.g. Hill-Davis [7]. In contrast to the Johnson transformation this approach does not rely on distributional assumptions.
- Note that first moments (from which we can calculate the reserves) w.r.t. Model 2.2 are rather sensitive on the structural parameters σ_j (c.f. (3.8)) whereas for Model 2.1 deviations in the estimation of ψ_j have only marginal influence (via posterior parameters, see Proposition 3.1 and (3.1)). Therefore we conclude that Model 2.1 is more stable for prediction.
- As an extension to the presented 4-parameter approximation within the Johnson family of distributions one could implement non-parametric approximations such as Edgeworth expansions or saddlepoint approximations based on higher moments. An essential assumption to apply such techniques is the existence of higher moments which is the case for Model 2.2 but for Model 2.1 higher moments may fail to exist. For an overview on this topic we refer to Butler [3].

Conclusion

We have studied the claims development result (CDR). The CDR is one of the major risk drivers in the profit and loss statement of a general insurance company and therefore needs a careful and appropriate risk assessment. In addition to the existing actuarial literature we determine higher order moments of the CDR distribution (for two different analytical models). The consideration of higher moments allows insurance companies to gain more insight into the shape of the distribution of the CDR, by e.g. considering the skewness and the excess kurtosis, which is essential for the application of techniques for estimating risk adjustments of future cash flows as discussed in IASB [9].

As proposed in Solvency II and applied in current practice, a proxy for the distribution of the CDR is obtained by fitting a shifted lognormal distribution to the estimates of the first two moments of the CDR. Within the 4-parameter Johnson family of distributions we refine the approximation for higher moments by including skewness and excess kurtosis parameters. The case study shows that the shifted lognormal approximation may result in too symmetric distributions which fail to fit the skewness and the excess kurtosis of the distribution of the CDR. Taking factors such as the shape of the distribution into account significantly influences the estimation of risk adjustments as discussed in IASB [9] and therefore knowledge about higher moments of the CDR contributes to a comprehensive assessment of the reserving risk.

References

[1] Bühlmann, H., De Felice, M., Gisler, A., Moriconi, F. and Wüthrich, M.V. (2009) Recursive credibility formula for chain ladder factors and the claims development result. *ASTIN Bulletin*, **39(1)**, 275-306.

[2] Bühlmann, H. and Gisler, A. (2005) *A Course in Credibility Theory and its Applications*. Springer.

[3] Butler, R.W. (2007) *Saddlepoint Approximations with Applications*. Cambridge University Press.

[4] Elderton, W.P. and Johnson, N.L. (1969) *System of Frequency Curves*. Cambridge University Press.

[5] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Springer.

[6] Gisler, A. and Wüthrich, M.V. (2008) Credibility for the chain ladder reserving method. *ASTIN Bulletin*, **38(2)**, 565-600.

[7] Hill, G.W. and Davis, A.W. (1968) Generalized asymptotic expansions of Cornish-Fisher type. *The Annals of Mathematical Statistics*, **39(4)**, 1264-1273.

[8] Hill, I.D., Hill, R. and Holder, R.L. (1976) Fitting Johnson curves by moments. *Applied Statistics*, **25(2)**, 180-189.

[9] IASB (2010) Exposure Draft Insurance Contracts. ED/2010/8, available under http://www.ifrs.org

[10] Johnson, N.L. (1965) Tables to facilitate fitting S_u frequency curves. *Biometrika*, **52(3/4)**, 547-558.

[11] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994) *Continuous Univariate Distributions* (Volume 1, 2nd Edition). Wiley series in probability and mathematical statistics.

[12] Mack, T. (1993) Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, **23(2)**, 213-225.

[13] MACK, T. (1994) *Measuring the variability of chain ladder reserve estimates.* Casualty Actuarial Society Forum 1, Spring 1994, 101-182.
[14] MCNEIL, A.J., FREY, R. and EMBRECHTS, P. (2005) *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.
[15] SALZMANN, R. and WÜTHRICH, M.V. (2010) Cost-of-capital margin for a general insurance liability runoff. *ASTIN Bulletin*, **40(2)**, 415-451.
[16] WÜTHRICH, M.V. and MERZ, M. (2008) *Stochastic Claims Reserving Methods in Insurance.* Wiley Finance.

ROBERT SALZMANN
ETH Zurich
RiskLab Switzerland
Department of Mathematics
8092 Zurich
Switzerland

MARIO V. WÜTHRICH
ETH Zurich
RiskLab Switzerland
Department of Mathematics
8092 Zurich
Switzerland

MICHAEL MERZ
University of Hamburg
Department of Business Administration
20146 Hamburg
Germany

A. APPENDIX

Proof of Proposition 3.1. We start with proving the result for Model 2.1.

(a) We denote the distribution of $C_{i,0}$ by π_i. By the chain rule of conditional probability and the Markov assumption in Model 2.1 the joint density of $(\mathbf{\Theta}, \mathcal{D}_{I+t})$ is given by

$$\pi\left(\mathbf{\Theta}, \mathcal{D}_{I+t}\right) = g(\mathcal{D}_{I+t-1}) \prod_{i+j \le I+t-1} \frac{\left(\theta_j \psi_j^{-2}\right)^{\psi_j^{-2}}}{\Gamma\left(\psi_j^{-2}\right)} (F_{i,j})^{\psi_j^{-2}-1} \exp\left(-\theta_j \psi_j^{-2} F_{i,j}\right)$$
$$\times \prod_{j=0}^{J-1} \frac{\left(f_j(\gamma_j - 1)\right)^{\gamma_j}}{\Gamma(\gamma_j)} \theta_j^{\gamma_j - 1} \exp\left(-f_j(\gamma_j-1)\theta_j\right) \prod_{i=0}^{I} \pi_i(C_{i,0}).$$

Note that $g(\mathcal{D}_{I+t-1})$ is some norming function depending on terms $C_{i,j} \in \mathcal{D}_{I+t-1}$. This implies that the posterior density of $\mathbf{\Theta}$, given $\mathcal{D}_{I+t}$, satisfies the following proportionality property

$$\pi\left(\mathbf{\Theta} \mid \mathcal{D}_{I+t}\right) \propto \prod_{j=0}^{J-1} \theta_j^{\frac{I-j+t}{\psi_j^2} + \gamma_j - 1} \exp\left(-\theta_j \left(\psi_j^{-2} \sum_{i=0}^{I-j-1+t} F_{i,j} + f_j(\gamma_j - 1)\right)\right).$$

These are independent gamma densities which proves the claim for Model 2.1.

(b) Similar as above for Model 2.2, the joint density of $(\boldsymbol{\mu}, \mathcal{D}_{I+t})$ is given as follows:

$$\pi(\boldsymbol{\mu}, \mathcal{D}_{I+t})$$
$$= g(\mathcal{D}_{I+t-1}) \prod_{i+j \le I+t-1} \frac{\exp\left(-\frac{(\log(F_{i,j}) - \mu_j)^2}{2\sigma_j^2}\right)}{\sigma_j \sqrt{2\pi}\, F_{i,j}} \prod_{j=0}^{J-1} \frac{\exp\left(-\frac{(\mu_j - \xi_j)^2}{2v_j^2}\right)}{v_j \sqrt{2\pi}} \prod_{i=0}^{I} \pi_i(C_{i,0}).$$

This implies that the posterior density of $\boldsymbol{\mu}$, given $\mathcal{D}_{I+t}$, satisfies the following proportionality property

$$\pi(\boldsymbol{\mu} \mid \mathcal{D}_{I+t}) \propto \prod_{j=0}^{J-1} \exp\left(-\frac{(\mu_j - \xi_{j,t})^2}{2v_{j,t}^2}\right),$$

with

$$\xi_{j,t} = \frac{\xi_j + v_j^2 \sigma_j^{-2} \sum_{i=0}^{I-j-1+t} \log(F_{i,j})}{1 + (I-j+t)\, v_j^2 \sigma_j^{-2}}, \text{ and } v_{j,t}^2 = \frac{v_j^2}{1 + (I-j+t)\, v_j^2 \sigma_j^{-2}}.$$

These are independent normal densities which proves the claim for Model 2.2. □

Proof of Proposition 3.6. Let us first prove the result for Model 2.1.

(a) Due to Proposition 3.1 and Corollary 3.3 we find for $n \in \mathbb{N}$ that

$$\left(\widehat{C}_{i,J}^{(1)}\right)^n = \left(C_{i,I-i+1}\right)^n \prod_{j=I-i+1}^{J-1} \left(\beta_j F_{I-j,j} + (1-\beta_j)\,\widehat{F}_j^{(0)}\right)^n.$$

With the binomial formula, this rewrites as

$$\left(\widehat{C}_{i,J}^{(1)}\right)^n = \left(C_{i,I-i+1}\right)^n \prod_{j=I-i+1}^{J-1} \sum_{l=0}^{n} \binom{n}{l} (\beta_j)^l \left(F_{I-j,j}\right)^l (1-\beta_j)^{n-l} \left(\widehat{F}_j^{(0)}\right)^{n-l}.$$

With this, we get

$$\begin{aligned}&\mathbb{E}\left[\left(\widehat{C}_{i,J}^{(1)}\right)^n \middle| \mathcal{D}_I\right]\\ &= \mathbb{E}\left[\left(C_{i,I-i+1}\right)^n \prod_{j=I-i+1}^{J-1} \sum_{l=0}^{n} \binom{n}{l} (\beta_j)^l \left(F_{I-j,j}\right)^l (1-\beta_j)^{n-l} \left(\widehat{F}_j^{(0)}\right)^{n-l} \middle| \mathcal{D}_I\right].\end{aligned}$$

Due to posterior independence the above expression on the right can be written as follows:

$$\left(C_{i,I-i}\right)^n \mathbb{E}\left[\left(F_{i,I-i}\right)^n \middle| \mathcal{D}_I\right] \prod_{j=I-i+1}^{J-1} \sum_{l=0}^{n} \binom{n}{l} (\beta_j)^l\, \mathbb{E}\left[\left(F_{I-j,j}\right)^l \middle| \mathcal{D}_I\right] (1-\beta_j)^{n-l} \left(\widehat{F}_j^{(0)}\right)^{n-l}.$$

Now we consider terms of the form

$$\mathbb{E}\left[\left(F_{I-j,j}\right)^k \middle| \mathcal{D}_I\right],$$

for $0 < k \leq n$ and $j = I - i, \ldots, J - 1$. By the assumptions of Model 2.1, the tower property for conditional expectations and Proposition 3.1 we get

$$\begin{aligned}\mathbb{E}\left[\left(F_{I-j,j}\right)^k \middle| \mathcal{D}_I\right] &= \mathbb{E}\left[\mathbb{E}\left[\left(F_{I-j,j}\right)^k \middle| \Theta, \mathcal{D}_I\right] \middle| \mathcal{D}_I\right] = \mathbb{E}\left[\theta_j^{-k} \middle| \mathcal{D}_I\right] \left(\psi_j^2\right)^k \frac{\Gamma\left(\psi_j^{-2} + k\right)}{\Gamma\left(\psi_j^{-2}\right)}\\ &= \left(\psi_j^2 c_{j,0}\right)^k \frac{\Gamma\left(\psi_j^{-2} + k\right)}{\Gamma\left(\psi_j^{-2}\right)} \frac{\Gamma\left(\gamma_{j,0} - k\right)}{\Gamma\left(\gamma_{j,0}\right)}.\end{aligned}$$

The above expression is obtained by applying twice formula (2.1) for the moments of a gamma distributed random variable. Since the last expression corresponds to $K_{k,j}$, this proves the proposition for Model 2.1.

(b) First we note that (c.f. (3.10))

$$\left(\widehat{C}_{i,J}^{(1)}\right)^n = \left(C_{i,I-i+1} \prod_{j=I-i+1}^{J-1} \exp\left(\frac{\sigma_j^2 + \nu_{j,1}^2}{2}\right) \exp\left(\widehat{Y}_j^{(1)}\right)\right)^n.$$

Hence we get

$$\mathbb{E}\left[\left(\widehat{C}_{i,J}^{(1)}\right)^n \middle| \mathcal{D}_I\right] = \mathbb{E}\left[\left(C_{i,I-i+1} \prod_{j=I-i+1}^{J-1} \exp\left(\frac{\sigma_j^2 + \nu_{j,1}^2}{2}\right) \exp\left(\widehat{Y}_j^{(1)}\right)\right)^n \middle| \mathcal{D}_I\right].$$

Due to posterior independence the above expression on the right can be written as follows:

$$(C_{i,I-i})^n \, \mathbb{E}\left[(F_{i,I-i})^n | \mathcal{D}_I\right] \prod_{j=I-i+1}^{J-1} \exp\left(n \frac{\sigma_j^2 + \nu_{j,1}^2}{2}\right) \mathbb{E}\left[\exp\left(n \widehat{Y}_j^{(1)}\right) \middle| \mathcal{D}_I\right].$$

By the assumptions of Model 2.2 and Proposition 3.1 we obtain

$$\begin{aligned}
\mathbb{E}\left[(F_{i,I-i})^n | \mathcal{D}_I\right] &= \mathbb{E}\left[\mathbb{E}\left[(F_{i,I-i})^n | \boldsymbol{\mu}, \mathcal{D}_I\right] | \mathcal{D}_I\right] = \mathbb{E}\left[\exp\left(n\mu_{I-i} + n^2 \frac{\sigma_{I-i}^2}{2}\right) \middle| \mathcal{D}_I\right] \\
&= \exp\left(\frac{n^2 \sigma_{I-i}^2}{2}\right) \exp\left(n \xi_{I-i,0} + \frac{n^2 \nu_{I-i,0}^2}{2}\right).
\end{aligned}$$

Furthermore we have that (see Corollary 3.5)

$$\mathbb{E}\left[\exp\left(n \widehat{Y}_j^{(1)}\right) \middle| \mathcal{D}_I\right] = \exp\left(n(1-\kappa_j) \widehat{Y}_j^{(0)}\right) \mathbb{E}\left[\exp(n\kappa_j Y_{I-j,j}) | \mathcal{D}_I\right].$$

Since $Y_{I-j,j}$, for given $\boldsymbol{\mu}$ and $\mathcal{D}_I$, are normally distributed with $\mathcal{N}(\mu_j, \sigma_j)$ and using Proposition 3.1 we find

$$\begin{aligned}
\mathbb{E}\left[\exp(n\kappa_j Y_{I-j,j}) | \mathcal{D}_I\right] &= \mathbb{E}\left[\mathbb{E}\left[\exp(n\kappa_j Y_{I-j,j}) \middle| \boldsymbol{\mu}, \mathcal{D}_I\right] \middle| \mathcal{D}_I\right] \\
&= \mathbb{E}\left[\exp\left(n\kappa_j \mu_j + \frac{(n\kappa_j)^2 \sigma_j^2}{2}\right) \middle| \mathcal{D}_I\right] \\
&= \exp\left(\frac{(n\kappa_j)^2 \sigma_j^2}{2}\right) \exp\left(n\kappa_j \xi_{j,0} + \frac{(n\kappa_j)^2 \nu_{j,0}^2}{2}\right).
\end{aligned}$$

Rearranging the terms accordingly completes the proof.

□

Proof of Proposition 3.7. For $i = I - J + 1$ we have that $\widehat{C}^{(1)}_{i,J} = C_{i,I-i+1}$. By the assumptions of Model 2.2 we find in this case that

$$\widehat{C}^{(1)}_{i,J}\Big|_{\boldsymbol{\mu}, C_{i,I-i}} \sim \mathcal{LN}\big(\log(C_{i,I-i}) + \mu_{I-i}, \sigma^2_{I-i}\big),$$

By formula (2.4) and Proposition 3.1 we obtain for $\widehat{C}^{(1)}_{i,J}$ and $i > I - j + 1$

$$\begin{aligned}
\widehat{C}^{(1)}_{i,J} &= \mathbb{E}[C_{i,J}|\mathcal{D}_{I+1}] \\
&= C_{i,I-i+1} \prod_{j=I-i+1}^{J-1} \mathbb{E}\big[\exp(\mu_j)|\mathcal{D}_{I+1}\big] \exp\left(\frac{\sigma_j^2}{2}\right) \\
&= C_{i,I-i+1} \prod_{j=I-i+1}^{J-1} \exp\left(\frac{\xi_j + \nu_j^2\sigma_j^{-2}\sum_{k=0}^{I-j-1+1}\log(F_{k,j})}{1+(I-j+1)\nu_j^2\sigma_j^{-2}}\right) \exp\left(\frac{\sigma_j^2}{2} + \frac{\nu_{j,1}^2}{2}\right) \\
&= F_{i,I-i} \prod_{j=I-i+1}^{J-1} \exp\left(\frac{\xi_j}{1+(I-j+1)\nu_j^2\sigma_j^{-2}}\right) \exp\left(\kappa_j \sum_{k=0}^{I-j} \log(F_{k,j})\right) \\
&\quad \times \exp\left(\frac{\sigma_j^2}{2} + \frac{\nu_{j,1}^2}{2}\right) C_{i,I-i}.
\end{aligned}$$

By collecting $\mathcal{D}_I$-measurable terms in b_j, i.e. ,

$$b_j = \exp\left(\kappa_j \sum_{k=0}^{I-j-1} \log(F_{k,j})\right) \exp\left(\frac{\xi_j}{1+(I-j+1)\nu_j^2\sigma_j^{-2}}\right) \exp\left(\frac{\sigma_j^2}{2} + \frac{\nu_{j,1}^2}{2}\right),$$

we get

$$\widehat{C}^{(1)}_{i,J} = F_{i,I-i} \prod_{j=I-i+1}^{J-1} \exp\Big(\kappa_j \log\big(F_{I-j,j}\big)\Big) C_{i,I-i}\, b_j.$$

We write

$$\widehat{C}^{(1)}_{i,J} = \exp\left(\log(F_{i,I-i}) + \sum_{j=I-i+1}^{J-1} \kappa_j \log\big(F_{I-j,j}\big)\right) \underbrace{C_{i,I-i} \prod_{j=I-i+1}^{J-1} b_j}_{=d_i'}.$$

Using the assumptions of Model 2.2 we obtain that

$$\log(F_{i,I-i}) + \sum_{j=I-i+1}^{J-1} \kappa_j \log\big(F_{I-j,j}\big)\Bigg|_{\boldsymbol{\mu},\mathcal{D}_I} \sim \mathcal{N}\left(\sum_{j=I-i}^{J-1} \kappa_j \mu_j, \sum_{j=I-i}^{J-1} \kappa_j^2 \sigma_j^2\right).$$

Hence we obtain

$$\widehat{C}_{i,J}^{(1)}\Big|_{\boldsymbol{\mu},\mathcal{D}_I} \sim \mathcal{LN}\left(\log(d_i') + \sum_{j=I-i}^{J-1} \kappa_j \mu_j, \sum_{j=I-i}^{J-1} \kappa_j^2 \sigma_j^2\right),$$

with κ_j given by (3.7) for $j > I - i$ and $\kappa_{I-i} = 1$. This completes the proof. □

Before we prove Theorem 3.8 we consider the following two lemmas.

Lemma A.1. *Assume Model 2.1. Choose $i > I - J$ and let $N_i = \min\{\gamma_{I-i,0}, \ldots, \gamma_{J-1,0}\}$. Then for $k, l \in \mathbb{N}$ with $k + l \leq N_i$ we obtain*

$$\mathbb{E}\left[F_{i,I-i}^k \left(\widehat{F}_{I-i}^{(1)}\right)^l \Big| \mathcal{D}_I\right] = \sum_{r=0}^{l} \binom{l}{r} \beta_{I-i}^r \left((1-\beta_{I-i})\widehat{F}_{I-i}^{(0)}\right)^{l-r} K_{k+r,I-i},$$

with $K_{m,j}$ as in Proposition 3.6 for $0 \leq m \leq n$. The above term is set equal to $K_{k,I-i}$ for $l = 0$.

Proof of Lemma A.1. Using Corollary 3.3 we get

$$\begin{aligned}\mathbb{E}\left[F_{i,I-i}^k \left(\widehat{F}_{I-i}^{(1)}\right)^l \Big| \mathcal{D}_I\right] &= \mathbb{E}\left[F_{i,I-i}^k \left(\beta_{I-i} F_{i,I-i} + (1-\beta_{I-i})\widehat{F}_{I-i}^{(0)}\right)^l \Big| \mathcal{D}_I\right] \\ &= \sum_{r=0}^{l} \binom{l}{r} \beta_{I-i}^r \left((1-\beta_{I-i})\widehat{F}_{I-i}^{(0)}\right)^{l-r} \mathbb{E}\left[F_{i,I-i}^{k+r} | \mathcal{D}_I\right].\end{aligned}$$

The moments of $F_{i,I-i}$ given $\mathcal{D}_I$ are already calculated in the proof of Proposition 3.6. □

Lemma A.2. *Assume Model 2.2 and choose $i > I - J + 1$, we obtain*

$$\begin{aligned}&\mathbb{E}\left[F_{i,I-i}^k \left(\widehat{F}_{I-i}^{(1)}\right)^l \Big| \mathcal{D}_I\right] \\ &= \exp\left((k+l)\xi_{I-i,0}\right) \exp\left(\frac{l}{2}\left(\sigma_{I-i}^2 + \nu_{I-i,1}^2\right) + \frac{1}{2}\left(k + \kappa_{I-i}\, l\right)^2 \left(\sigma_{I-i}^2 + \nu_{I-i,0}^2\right)\right).\end{aligned}$$

Proof of Lemma A.2. Using (3.10) we get

$$\begin{aligned}&\mathbb{E}\left[F_{i,I-i}^k \left(\widehat{F}_{I-i}^{(1)}\right)^l \Big| \mathcal{D}_I\right] \\ &= \left(\exp((1-\kappa_{I-i})\widehat{Y}_{I-i}^{(0)}) \exp\left(\frac{\sigma_{I-i}^2}{2} + \frac{\nu_{I-i,1}^2}{2}\right)\right)^l \mathbb{E}\left[\left(F_{i,I-i}\right)^{k+\kappa_{I-i} l} \Big| \mathcal{D}_I\right].\end{aligned}$$

From the assumptions of Model 2.2 and Proposition 3.1 we obtain for the expectation

$$\mathbb{E}[F_{i,j}^m|\mathcal{D}_I] = \exp\Big(\frac{1}{2}m^2\sigma_j^2\Big)\,\mathbb{E}[\exp(m\mu_j)|\mathcal{D}_I] = \exp\Big(m\xi_{j,0} + \frac{1}{2}m^2\Big(\sigma_j^2 + v_{j,0}^2\Big)\Big).$$

By $\widehat{Y}_{I-i}^{(0)} = \xi_{I-i,0}$, we obtain

$$\begin{aligned}\mathbb{E}\Big[F_{i,I-i}^k(\widehat{F}_{I-i}^{(1)})^l\Big|\mathcal{D}_I\Big] &= \left(\exp\Big((1-\kappa_{I-i})\widehat{Y}_{I-i}^{(0)}\Big)\exp\left(\frac{\sigma_{I-i}^2}{2}+\frac{v_{I-i,1}^2}{2}\right)\right)^l\\ &\quad\times \exp\Big(\big(k+\kappa_{I-i}\,l\big)\xi_{I-i,0}+\frac{1}{2}\big(k+\kappa_{I-i}\,l\big)^2\Big(\sigma_{I-i}^2+v_{I-i,0}^2\Big)\Big)\\ &= \exp(k\xi_{I-i,0})\exp\Big((1-\kappa_{I-i})\,l\,\widehat{Y}_{I-i}^{(0)}+\kappa_{I-i}\,l\,\xi_{I-i,0}\Big)\\ &\quad\times \exp\Big(\frac{l}{2}(\sigma_{I-i}^2+v_{I-i,1}^2)+\frac{1}{2}\big(k+\kappa_{I-i}\,l\big)^2\Big(\sigma_{I-i}^2+v_{I-i,0}^2\Big)\Big).\end{aligned}$$

□

Proof of Theorem 3.8. We consider

$$\mathbb{E}\Big[\Big(\widehat{C}_{\bullet,J}^{(1)}\Big)^n\Big|\mathcal{D}_I\Big] = \mathbb{E}\left[\left(\sum_{i=I-J+1}^{I}\widehat{C}_{i,J}^{(1)}\right)^n\middle|\mathcal{D}_I\right].$$

This generates a multinomial structure

$$\begin{aligned}&\mathbb{E}\left[\left(\sum_{i=I-J+1}^{I}\widehat{C}_{i,J}^{(1)}\right)^n\middle|\mathcal{D}_I\right]\\ &= \mathbb{E}\left[\sum_{\substack{k_0,\ldots,k_{J-1}\in\{0,\ldots,n\}\\ h_{J-1}=n}}\binom{n}{k_0,k_1,\ldots,k_{J-1}}\Big(\widehat{C}_{I-J+1,J}^{(1)}\Big)^{k_0}\ldots\Big(\widehat{C}_{i,J}^{(1)}\Big)^{k_{J-1}}\middle|\mathcal{D}_I\right],\end{aligned}$$

where

$$\binom{n}{k_0,k_1,\ldots,k_{J-1}} = \frac{n!}{k_0!k_1!\ldots k_{J-1}!}.$$

Therefore, we have to consider terms of the form

$$\mathbb{E}\Big[\Big(\widehat{C}_{I-J+1,J}^{(1)}\Big)^{k_0}\ldots\Big(\widehat{C}_{i,J}^{(1)}\Big)^{k_{J-1}}\Big|\mathcal{D}_I\Big].$$

By the independence of accident years, formula (3.1) and the definition of $\widehat{F}_j^{(1)}$ (c.f. Corollary 3.2), we see that $C_{i,I-i+1}$, $\widehat{F}_{I-i+1}^{(1)}, \ldots \widehat{F}_{J-1}^{(1)}$ are independent, given $\mathcal{D}_I$. Hence,

$$\mathbb{E}\left[\left(\widehat{C}_{I-J+1,J}^{(1)}\right)^{k_0} \ldots \left(\widehat{C}_{i,J}^{(1)}\right)^{k_{J-1}} \middle| \mathcal{D}_I\right] = \prod_{s=0}^{J-1}\left(\mathbb{E}\left[C_{I-J+1+s,J-s}^{k_s}\left(\widehat{F}_{J-1-s}^{(1)}\right)^{(n-\sum_{n=0}^{s} k_n)} \middle| \mathcal{D}_I\right]\right).$$

(a) First we prove the result for Model 2.1. Having decoupled the problem into independent problems (similar to Salzmann-Wüthrich [15], proof of Theorem 4.9), we are left with the calculation of terms of the form

$$\mathbb{E}\left[F_{i,I-i}^{k}\left(\widehat{F}_{I-i}^{(1)}\right)^{l} \middle| \mathcal{D}_I\right],$$

for $k + l \leq n < N_i$ with $N_i = \min\{\gamma_{I-i,0}, \ldots, \gamma_{J-1,0}\}$ for all $i \in \{I-J+1, \ldots, I\}$ which is done in Lemma A.1. In summary for $n < \min\{\gamma_{0,0}, \ldots, \gamma_{J-1,0}\}$ we obtain

$$\mathbb{E}\left[\left(\sum_{i=I-J+1}^{I} \widehat{C}_{i,J}^{(1)}\right)^{n} \middle| \mathcal{D}_I\right] = \sum_{\substack{k_0,\ldots,k_{J-1}\in\{0,\ldots,n\} \\ h_{J-1}=n}} \binom{n}{k_0, k_1, \ldots, k_{J-1}} \prod_{s=0}^{J-1}\Bigg(C_{I-J+1+s,J-1-s}^{k_s}$$
$$\times \sum_{r=0}^{n-h_s}\binom{n-h_s}{r}\beta_{J-1-s}^{r}\left((1-\beta_{J-1-s})\,\widehat{F}_{J-1-s}^{(0)}\right)^{n-h_s-r} K_{k_s+r,J-1-s}\Bigg),$$

where the last sum is set equal to $K_{k_s,J-1-s}$ if $n - h_s = 0$.

(b) Finally, we show the result for Model 2.2.

With Lemma A.2 we obtain

$$\mathbb{E}\left[\left(\sum_{i=I-J+1}^{I} \widehat{C}_{i,J}^{(1)}\right)^{n} \middle| \mathcal{D}_I\right] = \sum_{\substack{k_0,\ldots,k_{J-1}\in\{0,\ldots,n\} \\ h_{J-1}=n}} \binom{n}{k_0, k_1, \ldots, k_{J-1}} \prod_{s=0}^{J-1}\Bigg(C_{I-J+1+s,J-1-s}^{k_s}$$
$$\times \exp\left((k_s + (n-h_s))\,\xi_{J-1-s,0}\right)$$
$$\times \exp\Bigg(\frac{(k_s + \kappa_{J-1-s}(n-h_s))^2}{2}\left(\sigma_{J-1-s}^2 + \nu_{J-1-s,0}^2\right)$$
$$+ \frac{(n-h_s)}{2}\left(\sigma_{J-1-s}^2 + \nu_{J-1-s,1}^2\right)\Bigg)\Bigg)$$

$\square$